

Repensando a Web Semântica

João Sávio Ceregatti Longo

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)

Campinas – SP – Brazil

joaosavio@gmail.com

Abstract. *The Semantic Web is a Tim Berners-Lee's fantastic idea, but not yet consolidated. Most authors say that its failure is related to many reasons, but mainly to its complexity. In this paper, first it will be seen various criticisms, and then, it will be presented some alternatives that promise to solve the problems mentioned previously.*

Resumo. *A Web Semântica é uma ideia fantástica de Tim Berners-Lee, mas que ainda não se consolidou. A maioria dos autores afirma que o seu atual fracasso está relacionado a diversos fatores, mas principalmente por sua complexidade. Neste artigo, primeiramente serão vistas várias críticas e, em seguida, serão apresentadas algumas alternativas que prometem resolver os problemas anteriormente citados.*

1. Introdução

A Web Semântica será uma extensão da Web atual, em que as informações não serão apenas mostradas, mas sim interpretadas, modificadas e processadas pelas máquinas. Existirão agentes que farão tarefas para o usuário automaticamente, tais como marcar uma consulta médica ou fazer compras. Mas, isto ainda não é possível, pois o conteúdo da Web está formatado apenas para humanos entenderem e, por isso, os computadores apenas mostram as informações às pessoas [Berners-Lee et al, 2001].

Assim, começaram a surgir ideias para que os dados fossem ligados a outros e pudessem ser interpretados pelos computadores. Neste contexto, pesquisadores desenvolveram soluções que resolveriam este problema, como os formatos RDF (*Resource Description Format*), RDF Schema e OWL (*Web Ontology Language*), pois o XML (*eXtensible Markup Language*) sozinho não era suficiente.

O resultado disso foi que, infelizmente, a comunidade de usuários, desenvolvedores e empresas não aderiu ainda à Web Semântica, por vários motivos que serão citados neste artigo. Vários autores fizeram críticas muito pertinentes às atuais soluções apresentadas, e propõem abordagens não tão radicais, pelo menos por enquanto, que fariam com que a Web Semântica seja incorporada à Web atual de uma forma mais natural.

2. Críticas

Hendler (2007) acredita que existe o chamado “lado negro da Web Semântica”, fazendo alusão não a Darth Vader e o lado negro da força, mas sim ao lado negro da Lua, referindo-se à falta de percepção de alguns pesquisadores da área de Inteligência Artificial quanto à aspectos sociais. A Web Semântica prega a introdução de *tags* que também descrevam a informação às máquinas, e para isso, faz uso de metadados (em

outras palavras, dados sobre dados). Estes, por sua vez, ainda não são suficientes, pois eles deverão ser ligados à ontologias, que definirão os termos e suas relações. Ainda, uma ontologia tem uma taxionomia e um conjunto de regras de inferência [Frauenfelder, 2001]. Em resumo, eles querem fazer algo tão complexo que propõe uma modificação muito radical, não criando comunidades fortes, tanto em quantidade, quanto em qualidade [McCool, 2005].

Algo complicado é muito mais demorado a ser produzido. Logo, as empresas ainda não aderiram à Web Semântica, pois necessitam de um *feedback* rápido dos clientes [Genadinik, 2010]. Existem mais problemas que afastam as companhias, como a não maturidade da Web Semântica em relação à padrões como HTTP (*HyperText Transfer Protocol*) e HTML. Também, o conhecimento necessário para a criação de aplicações que usem os aspectos semânticos requerem que os desenvolvedores tenham um profundo entendimento teórico do assunto, e os padrões da Web Semântica não têm muita similaridade com o que é usado atualmente [Dokulil et al, 2008]. Assim, falta mão-de-obra qualificada, sendo praticamente impossível achar um lingüista computacional. E, caso seja encontrado, e este profissional sair da empresa, as chances de sobrevivência de um projeto são muito pequenas [Genadinik, 2010].

McCool (2005) cita que os formatos RDF, RDF Schema e OWL tem tido taxas de utilização extremamente baixas e os projetos existentes têm problemas de cobertura qualidade, apresentando vários erros, como por exemplo, o TAP [2005], Swoogle [2007], SIMILE [2008] e KnowItAll [2008]. O autor - que participou do desenvolvimento do projeto TAP - ainda diz que “sem uma radical simplificação, a Web Semântica continuará com participação limitada e com poucas aplicações convincentes”. Com isso, fica uma pergunta pertinente: como utilizar algo se os próprios exemplos não convencem?

Outro fator são as promessas da Web Semântica de que alguém poderia marcar uma consulta média baseado em um serviço de classificação confiável. De fato, com a tecnologia que temos hoje, ainda não dá para confiar totalmente nas máquinas. Assim, conversar com um amigo ou participar de uma boa comunidade é mais confiável que acreditar no que uma máquina diz [McCool, 2005]. Ainda não dá para se implantar serviços deste tipo, pois esta parte da confiança ainda precisa ser desenvolvida.

Finalizando, com a utilização de ontologias, começam a surgir alguns problemas técnicos. O próprio Ramanathan V. Guha - que conduziu o desenvolvimento do RDF - admitiu que existem alguns problemas realmente grandes que necessitam ser resolvidos primeiro, referindo-se à adoção Web Semântica proposta. Por exemplo, existem muitas preocupações com performance, pois quando regras são adicionadas a um sistema, o tempo para tirar conclusões de dados, metadados e ontologias cresce muito rapidamente [Frauenfelder, 2001]. Na Web Semântica de Tim Berners-Lee, existirão inúmeras ontologias, então há grande possibilidade da lentidão tomar conta da Web.

3. Algumas alternativas

É de consenso de vários autores que a Web Semântica deveria ser mais simples, para se aproximar de usuários, empresas e desenvolvedores, deixando assim de ser apenas algo idealizado. Podemos observar como a simplicidade leva pessoas a usarem as aplicações citando a Wikipedia - a enciclopédia na internet - que suporta sinônimos, desambiguação e categorias, ou seja, um conjunto simples de semântica. Quanto a isso, a frase de Hendler (2007) ilustra bem: “um pouquinho de semântica, multiplicada por bilhões de coisas que possam ser aplicadas na Web, pode ter muito poder”.

A ideia da Web Semântica é muito boa para ser descartada. Mas, é preciso simplificá-la para que as comunidades se fortaleçam e seu uso seja iminente. Logo, muitos pesquisadores têm se concentrado em uma evolução mais lenta da Web Semântica [Dokulil et al, 2008].

A seguir, veremos algumas abordagens que têm como objetivo inserir semântica na Web, mas de uma forma mais sutil.

Abordagem NEW

Existe uma abordagem chamada NEW (*Named Entity Web*) proposta por Rob McCool [2006], que removeria as classes, relações e triplas da Web Semântica. A principal ideia é adicionar parâmetros à *tags* já existentes na marcação HTML, o que já melhoraria muito para buscadores de páginas, pois conceitos estariam presentes nas páginas. Assim, esta abordagem tem como objetivo ser apenas uma extensão da Web atual e não precisa de um formalismo tão forte quanto ontologias para funcionar. “De fato, isso é vital para que o trabalho pesado seja removido em favor da participação”, cita McCool (2006).

Como exemplo, considere uma página pessoal (Figura 1) e outra de um dado laboratório de pesquisa (Figura 2). Podemos ver que o estudante é membro do laboratório que, por sua vez, referencia o aluno como membro. Percebe-se também a presença da *tag* `<meta>` identificada como *entity-name* ou *entity-type*, que provê metadados sobre a página, e no caso, indica sobre o que é a página, atribuindo-a uma entidade com nome e tipo.

```
<html>
<title>Joao Savio Home Page</title>
<meta name="entity-name" content="Joao Savio">
<meta name="entity-type" content="person">


<p>
  <a rel="entity-name-preferred">Joao Savio</a> is a <a rel="entity-type">student</a> at
  <a rel="entity-ident" href="http://www.unicamp.br/">Unicamp</a>'s <a rel="entity-ident"
  href="https://www.lis.ic.unicamp.br/">Laboratory of Information Systems</a>.
</p>
</html>
```

Figura 1 - Exemplo da abordagem NEW

```

<html>
<title> Laboratory of Information Systems Home Page</title>
<meta name="entity-name-preferred" content="Unicamp's Laboratory of Information Systems ">
<meta name="entity-name" content="Laboratory of Information Systems ">
<meta name="entity-type" content="organization">
<p>
  <a rel="entity-name">LIS</a> conducts research in the areas of <a rel="entity-ident">Scientific
  Databases </a> and <a rel="entity-ident">Geographic Information Systems </a> in the <a rel="entity-
  ident" href="http://www.ic.unicamp.br/">Institute of Computing</a> at <a rel="entity-ident" href=
  "http://www.unicamp.br/">Unicamp</a>.
</p>

<p>
  LIS people:
</p>
<ul>
  <li>
    <a rel="entity-ident" href="http://joaosavio.wordpress.com">Joao Savio</a>
  </li>
  ...
</ul>
</html>

```

Figura 2 - Exemplo da abordagem NEW (cont.)

Este novo esquema abre a possibilidade para algumas consultas semânticas. Por exemplo, a página do aluno (Figura 1) poderia ser facilmente encontrada procurando-se por “pessoas que fazem pesquisa em Sistema de Informação Geográfica”. Isto já seria uma grande evolução.

Microformatos

Microformatos são marcações que visam a colocar semântica nas páginas HTML [Yu, 2011]. Um microformato é composto por um conjunto de outros microformatos, formando um vocabulário, que cresce de acordo com a necessidade.

Como exemplo, podemos citar o *hCard* e *hCalendar*. O primeiro é utilizado para especificar informações básicas de pessoas e organizações, como nome, endereço, URLs, localização geográfica, etc, e o segundo objetiva descrever eventos em uma página HTML.

Eles possuem uma curva de aprendizado pequena, pois utilizam o máximo possível das próprias *tags* HTML.

RDFa

Atualmente, a W3C (*World Wide Web Consortium*) está trabalhando no desenvolvimento do RDFa (*Resource Description Framework in attributes*) que é uma sintaxe para expressar dados estruturados em qualquer linguagem de marcação [Yu, 2011]. Já existe uma versão para XHTML (*eXtensible Hypertext Markup Language*), mas em breve estará disponível também para HTML.

Possui atributos que incorporam a grande maioria dos grafos RDF [Adidas, 2008], ou seja, é uma maneira simples de expressar as triplas RDF dentro das páginas Web. Segundo Yu (2011), como existem vários jeitos de expressar RDF, a utilização do

RDFa forçará tanto humanos quanto máquinas a utilizarem somente uma maneira, e isto é uma das maiores motivações para sua adoção.

Em comparação aos microformatos, ele é um padrão W3C e possui uma sintaxe independente de vocabulário, pois a marcação semântica é feita baseada em um novo conjunto de atributos.

GRDDL

Outro padrão que está sendo desenvolvido pela W3C é o GRDDL (*Gleaning Resource Descriptions from Dialects of Languages*), que é um formato de marcação que permite facilmente a extração de triplas RDF de documentos XML [W3C, 2007]. Basicamente, ele introduz marcações para declarar a compatibilidade com RDF e ligar o documento a algoritmos extratores [Yu, 2011].

Uma de suas utilizações será para extrair RDF de documentos HTML anotados semanticamente com RDFa ou microformatos [Adidas, 2008].

HTML5

A especificação da próxima geração da linguagem de marcação HTML, que ficará pronta em 2014 [Shankland, 2011], promete a incorporação de um pouco de semântica. Será introduzido no HTML5 algumas novas *tags* que realmente impulsionarão práticas de otimização de motores de busca [Weishaupt, 2010].

As novas *tags* `<article>` e `<section>` ajudarão os buscadores a separar o conteúdo relevante de uma página do restante, acabando com a bagunça criada pela *tag* `<div>`. Também, o elemento `<nav>` identificará um grupo de *links*, representando a seção de navegação de um site, e em `<aside>` estarão os conteúdos não relevantes a buscas, como anúncios e *links* de patrocinadores. A Figura 3 representa o novo esquema de uma página utilizando HTML5.



Figura 3 - Estrutura de uma página utilizando HTML5 [Wright, 2009]

4. Conclusão

Existem críticas muito duras e pertinentes em relação à Web Semântica, inclusive algumas que fariam muitos pensarem em descartar esta ideia, pois é uma ambição muito grande. Por outro lado, deve-se considerar que a adição de semântica na Web possibilitará a extração de informações relevantes das páginas, o que permitiria utilizá-las em aplicações e ferramentas, sendo algo muito poderoso. Daria-se mais poder às máquinas, que por sua vez, tornariam a experiência dos usuários ainda mais interessante.

Foi necessário o primeiro passo de Tim Berners-Lee, para imaginar o quão poderosa a Web pode se tornar e assim caminhar nesta direção. Contudo, percebeu-se que propondo modificações tão radicais para a Web de hoje não surtiu efeito, pois o aspecto social foi esquecido. Então, fazer com que os conceitos da Web Semântica sejam incorporados pouco a pouco pela comunidade de usuários, desenvolvedores e empresas, beneficiará a todos.

Portanto, muita coisa ainda precisa ser melhorada até que a Web esteja no ponto que Tim Berners-Lee quer. Mas, enquanto isso, a cada pequena parte da ideia original que for incorporada, fará com que se chegue cada vez mais perto da tão sonhada Web Semântica.

5. Referências

- ADIDAS, B. hGRDDL: Bridging microformats and RDFa. *Web Semantics: Science, Services and Agents on the World Wide Web*, v.6, no. 1, Semantic Web and Web 2.0, Fevereiro/2008. Disponível em <<http://www.sciencedirect.com/science/article/B758F-4R53WD4-5/2/be46f87811ae4142830741cf502d009d>>. Acesso em 22/03/2011;
- DOKULIL, J., et al. Everything You Ever Wanted to Learn from the Semantic Web but Were Unable to Ask. *Advanced Engineering Computing and Applications in Sciences*, 2008. ADVCOMP '08. The Second International Conference on Advanced Engineering Computing and Applications in Sciences, pp. 21-26, Setembro-Outubro/2008. Disponível em <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4640987>>. Acesso em 17/03/2011;
- BERNERS-LEE, T., et al. The Semantic Web. *Scientific American*, Maio/2001. Disponível em <<http://www.dblab.ntua.gr/~bikakis/SW.pdf>>. Acesso em 17/03/2011;
- FRAUENFELDER, M. A Smarter Web. *Technology review*, Novembro/2001. Disponível em <http://ai.kaist.ac.kr/~sjcho/semantic-web/A_Smarter_Web.pdf>. Acesso em 15/03/2011;
- GENADINIK, A. Top 10 Reasons Not To Do a Semantic Web Project. 2010. Disponível em <<http://www.devx.com/semantic/Article/44832>>. Acesso em 16/03/2011;

- HENDLER, J. The Dark Side Of Semantic Web. *Intelligent Systems, IEEE*, vol.22, no.1, p.2-4, Jan.-Feb. 2007. Disponível em <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4078947>>. Acesso em 16/03/2011;
- MCCOOL, R. Rethinking the Semantic Web, Part I. *Internet Computing, IEEE*, v.9, no.6, pp. 88, 86-87, Novembro-Dezembro/2005 . Disponível em <<http://dx.doi.org/10.1109/MIC.2005.133>>. Acesso em 15/03/2011;
- MCCOOL, R. Rethinking the Semantic Web, Part II. *Internet Computing, IEEE*, v.10, no.1, pp. 93-96, Janeiro-Fevereiro/2006. Disponível em <<http://dx.doi.org/10.1109/MIC.2006.18>>. Acesso em 15/03/2011;
- SHANKLAND, S. HTML5 spec set for 2014 completion. 2011. Disponível em <http://news.cnet.com/8301-30685_3-20031673-264.html?tag=topStories1>. Acesso em 24/03/2011;
- WEISHAUP, J. HTML 5 & RDFa: Unlocking the Semantic Web. 2010. Disponível em <<http://www.joergweishaupt.com/seo/design-and-usability/html5-rdfa-unlocking-the-semantic-web.html>>. Acesso em 20/03/2011;
- WRIGHT, T. HTML 5 and The Future of Web. Smashing Magazine, Julho/2009. Disponível em <<http://www.smashingmagazine.com/2009/07/16/html5-and-the-future-of-the-web/>>. Acesso em 21/03/2011;
- W3C. HTML 4.01 Specification. 1999. Disponível em <<http://www.w3.org/TR/html401/>>. Acesso em 23/03/2011;
- W3C. GRDDL Primer. 2007. Disponível em <<http://www.w3.org/TR/grddl-primer/>>. Acesso em 23/03/2011;
- W3C. HTML+RDFa 1.1. 2011. Disponível em <<http://dev.w3.org/html5/rdfa/>>. Acesso em 20/03/2011;
- YU, L. A Developer's Guide to the Semantic Web, Capítulo 3 - Other RDF-Related Technologies: Microformats, RDFa, and GRDDL. 2011. Editora Springer Berlin Heidelberg. Disponível em <<http://www.springerlink.com/content/j0t725m615713977/>>. Acesso em 22/03/2011;
- Swoogle Semantic Web Search and Metadata Engine. 2007. Disponível em <<http://swoogle.umbc.edu>>. Acesso em 17/03/2011;
- TAP - The Alpiri Project. 2005. Disponível em <<http://ksl.stanford.edu/projects/TAP>>. Acesso em 17/03/2011;
- SIMILE - Semantic Interoperability of Metadata and Information in unLike Environments. 2008. Disponível em <<http://simile.mit.edu>>. Acesso em 17/03/2011;
- KnowItAll. 2008. Disponível em <www.cs.washington.edu/research/knowitall>. Acesso em 17/03/2011.